Continuous Logic and Learning Bounds

Aaron Anderson (joint with Michael Benedikt)

UPenn

September 27, 2025

Probably Approximately Correct Learning

A class $\mathcal{H}$ of functions $X \to [0,1]$ is PAC learnable when

- for every $\varepsilon, \delta > 0$, there is n such that,

Probably Approximately Correct Learning

A class $\mathcal{H}$ of functions $X \to [0, 1]$ is PAC learnable when

- for every $\varepsilon, \delta > 0$, there is n such that,
- given i.i.d. random $(x_1, y_1), \ldots, (x_n, y_n) \in X \times [0, 1]$,

Probably Approximately Correct Learning

A class $\mathcal{H}$ of functions $X \to [0,1]$ is PAC learnable when

- for every $\varepsilon, \delta > 0$, there is n such that,
- given i.i.d. random $(x_1, y_1), \ldots, (x_n, y_n) \in X \times [0,1]$,
- we can choose $h \in \mathcal{H}$ (with $h(x_i) \approx y_i$), such that

Probably Approximately Correct Learning

A class $\mathcal{H}$ of functions $X \to [0,1]$ is PAC learnable when

- for every $\varepsilon, \delta > 0$, there is n such that,
- given i.i.d. random $(x_1, y_1), \ldots, (x_n, y_n) \in X \times [0,1]$,
- we can choose $h \in \mathcal{H}$ (with $h(x_i) \approx y_i$), such that
- with probability at least $1 - \delta$,

Probably Approximately Correct Learning

A class $\mathcal{H}$ of functions $X \to [0,1]$ is PAC learnable when

- for every $\varepsilon, \delta > 0$, there is n such that,
- given i.i.d. random $(x_1, y_1), \ldots, (x_n, y_n) \in X \times [0,1]$,
- we can choose $h \in \mathcal{H}$ (with $h(x_i) \approx y_i$), such that
- with probability at least $1 - \delta$,
- $\mathbb{E}\left[|h(x_{n+1}) - y_{n+1}|\right]$ is within ε of the best case for all $h \in \mathcal{H}$.

Model Theory to Learnability

- Let M be a structure, $\phi(x; y)$ a formula.
- Define $\mathcal{H} = \{\phi(x; b) : b \in M^y\}$, where $\phi(x; b) : M^x \to \{0, 1\}$.

Model Theory to Learnability

- Let M be a structure, $\phi(x; y)$ a formula.
- Define $\mathcal{H} = \{\phi(x; b) : b \in M^y\}$, where $\phi(x; b) : M^x \to \{0, 1\}$.

Fact

The properties in each row are equivalent:

Model Theory	*Combinatorics*	*Learning Theory*
ϕ *is NIP*	*finite VC dimension*	*PAC learnable*
ϕ *is stable*	*finite Littlestone dimension*	*online learnable*

Model Theory to Learnability

- Let M be a structure, $\phi(x; y)$ a formula.
- Define $\mathcal{H} = \{\phi(x; b) : b \in M^y\}$, where $\phi(x; b) : M^x \to \{0, 1\}$.

Fact

The properties in each row are equivalent:

Model Theory	*Combinatorics*	*Learning Theory*
ϕ *is NIP*	*finite VC dimension*	*PAC learnable*
ϕ *is stable*	*finite Littlestone dimension*	*online learnable*

- This gives us many learnable classes of $\{0, 1\}$-valued functions.

Continuous Logic

- *Continuous logic* lets $\phi(x; y)$ define $[0, 1]$-valued functions.
- The properties from the table generalize, and remain equivalent:

Continuous Logic

- *Continuous logic* lets $\phi(x; y)$ define $[0, 1]$-valued functions.
- The properties from the table generalize, and remain equivalent:

Theorem (A., Benedikt)

Stability in continuous logic is equivalent to online learnability.

Continuous Logic

- *Continuous logic* lets $\phi(x; y)$ define $[0, 1]$-valued functions.
- The properties from the table generalize, and remain equivalent:

Theorem (A., Benedikt)

Stability in continuous logic is equivalent to online learnability.

- To construct some continuous logic formulas, use *randomization*...

The Randomization

Definition

The *randomization* of $\phi(x; y)$ is the class of $[0,1]$-valued functions

- on the set of random variables $\mathbf{x}$ on M^x
- indexed by random variables $\mathbf{y}$ on M^y
- defined by

$$\mathbf{x} \mapsto \mathbb{P}[\phi(\mathbf{x}, \mathbf{y})].$$

The Randomization

Definition

The *randomization* of $\phi(x; y)$ is the class of $[0,1]$-valued functions

- on the set of random variables $\mathbf{x}$ on M^x
- indexed by random variables $\mathbf{y}$ on M^y
- defined by
$$\mathbf{x} \mapsto \mathbb{P}[\phi(\mathbf{x}, \mathbf{y})].$$

Fact (Ben Yaacov, Keisler)

Randomization preserves NIP, stability (in continuous logic).

The Randomization

Definition

The *randomization* of $\phi(x; y)$ is the class of $[0, 1]$-valued functions

- on the set of random variables $\mathbf{x}$ on M^x
- indexed by random variables $\mathbf{y}$ on M^y
- defined by

$$\mathbf{x} \mapsto \mathbb{P}[\phi(\mathbf{x}, \mathbf{y})].$$

Fact (Ben Yaacov, Keisler)

Randomization preserves NIP, stability (in continuous logic).

Theorem (A., Benedikt)

Randomization preserves PAC/online learnability, with explicit bounds.

Thank you, MAMLS!

Generalizing VC Dimension to Continuous Logic

NIP in continuous logic means finite γ-fat-shattering dimension for every $\gamma > 0$:

Definition

Let $\mathcal{H}$ be a class of functions $X \to [0,1]$ and let $\gamma > 0$. We say $\mathcal{H}$ has *γ-fat-shattering dimension* at least n when there are

- $x_1, \ldots, x_n \in X$
- functions $h_E \in \mathcal{H}$ for each $E \subseteq \{1, \ldots, n\}$ satisfying

$$h_{E_0}(x_i) + \gamma \leq h_{E_1}(x_i)$$

  whenever $x_i \notin E_0, x_i \in E_1$.

Sequential fat-shattering dimension replaces subsets $E \subseteq \{1, \ldots, n\}$ with branches of a binary tree of depth n.